International Journal of Engineering Research
Volume No.3, Issue No.3, pp : 177-181
ISSN:2319-6890)(online),2347-5013(print)
01 March 2014

A study of various Fuzzy Clustering Algorithms

Nidhi Grover
*Assistant Professor, Institute of Information Technology & Management, Delhi, India
nidhigrover040189@gmail.com

Abstract : ***In data mining clustering techniques are used to group together the objects showing similar characteristics within the same cluster and the objects demonstrating different characteristics are grouped into clusters. Clustering approaches can be classified into two categories namely- Hard clustering and Soft clustering. In hard clustering data is divided into clusters in such a way that each data item belongs to a single cluster only while soft clustering also known as fuzzy clustering forms clusters such that data elements can belong to more than one cluster based on their membership levels which indicate the degree to which the data elements belong to the different clusters. Fuzzy C-Means is one of the most popular fuzzy clustering techniques and is more efficient that conventional clustering algorithms. In this paper we present a study on various fuzzy clustering algorithms such as Fuzzy C-Means Algorithm (FCM), Possibilistic C-Means Algorithm (PCM), Fuzzy Possibilistic C-Means Algorithm (FPCM) and Possibilistic Fuzzy C Means Algorithm (PFCM) with their respective advantages and drawbacks.***

Keywords : Hierarchical Clustering, Partitional Clustering, Soft clustering, Hard clustering, FCM, PCM, FPCM, PFCM

I. Introduction

Knowledge discovery in databases or KDD process refers to the overall procedure of discovering useful knowledge from data. It involves the evaluation and possible interpretation of the patterns in order to extract knowledge. KDD has evolved from interaction among various fields such as artificial Intelligence, machine learning, pattern recognition, database, statistics, knowledge representation for intelligent systems etc. A specific step in KDD process known as data mining, deals with applying various algorithms for extracting useful patterns and knowledge from raw data without the additional steps of the KDD process. Data mining involves analyzing observational datasets, finding out unsuspected relationships among them and summarizing the data in a clear, useful and understandable way for the data users [1]. Clustering techniques are used in data mining to group similar objects into the same classes whereas the objects showing different characteristics are grouped in different classes [2]. Cluster analysis has applications in different fields such as data mining, geographical data processing, medicine, classification of statistical findings in social studies and so on. These fields deal with huge amounts of data so the techniques required for handling such enormous amounts of data be efficient both in terms of the number of data set scans and memory usage [1]. Clustering is used for breaking data into related components so that patterns and order becomes visible. Large volumes of data are examined thoroughly to extract useful information in the form of new relationships, patterns, or clusters, for the purpose of decision-making by a user [2]. Clustering is different from classification since it deals with unsupervised learning of unlabeled data so, clustering algorithms can be safely used on a data set without much knowledge of it while in classification the class-prediction is done on unlabeled data after a supervised learning on pre-labeled data [3]. A cluster is usually represented as grouping of similar data points around a center known as centroid or it may be defined as prototype data instance nearest to the centroid. A cluster can be represented with or without a well-defined boundary such that those clusters with well-defined boundaries are called crisp clusters whereas those without a well-defined boundary are called fuzzy clusters. Clustering combines observed objects into clusters satisfying the main criteria described as follows [4]:

- Objects belonging to the same cluster are similar to each other i.e. each cluster is homogeneous.
- Each cluster should be different from other clusters such that objects belonging to one cluster are different from the objects present in other clusters i.e. different clusters are non-homogenous.

Clustering technique provides many advantages but the two most important benefits of clustering can be outlined as follows [3]:

1. Detection and handling of noisy data and outliers is relatively easier.
2. It provides the ability to deal with the data having different types of variables such as continuous variable that requires standardized data, binary variable, nominal variable, ordinal variable and mixed variables.

II. Literature Review

Data clustering is the process of dividing data elements into groups or clusters such that items in the same class are similar and items belonging to different classes are dissimilar. Different measures of similarity such as distance, connectivity, and intensity may be used to place different items into clusters. The similarity measure controls how the clusters are formed and depends on the nature of the data and the purpose of clustering data. Clustering technique can be hard or soft Clustering techniques can be classified into supervised clustering that demands human interaction to decide the clustering criteria and unsupervised clustering that decides the clustering criteria itself [2]. The two types of classic clustering techniques are defined as follows:

- Hierarchical Clustering Techniques
- Partitional Clustering Techniques

Hierarchical Clustering

The Hierarchical techniques produce a nested sequence of partition, with an inclusive single cluster at the top and single clusters of individual points at the bottom where each

International Journal of Engineering Research
Volume No.3, Issue No.3, pp : 177-181

ISSN:2319-6890)(online),2347-5013(print)
01 March 2014

intermediate level is regarded as combining two clusters from the next lower level or splitting a cluster from the next higher level [5]. Hierarchical algorithms create a hierarchical decomposition of the objects and are either agglomerative (bottom-up) or divisive (top-down) [2]. The dendogram is a tree like structure that graphically portrays the result of hierarchical clustering algorithm and displays the merging process and the intermediate clusters [5]. The dendogram at the right shows how four points can be merged into a single cluster. For document clustering, this dendogram provides a hierarchical index.

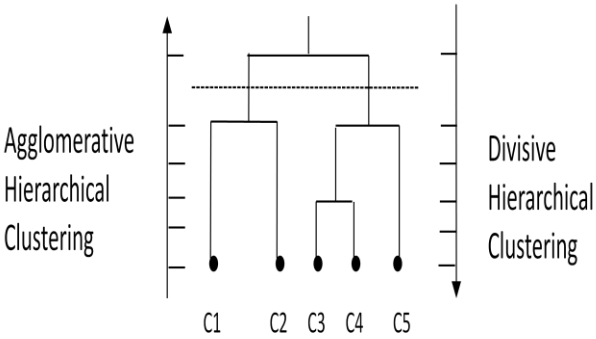

Fig 1: Agglomerative and Divisive Clustering

Hierarchical clustering has many variations. The top-down procedure in hierarchical clustering is called divisive hierarchical clustering. Here we start with one top most cluster and split the cluster at each step until only singleton clusters of individual points remain. The decision to be taken is which cluster to split and how to perform the split [5, 6]. On the other hand if the procedure starts at the bottom to form a pair of the data points with the shortest distance between them and merge the pair into a common representative and iterates till only one cluster representing all the original data points remain then this bottom-up procedure is called agglomerative such that the starting points are the individual clusters and, at each step, merge the most similar or closest pair of clusters [5, 6].

Partitional Clustering

The partitioning clustering method partitions a collection of elements into a set of non-overlapping and un-nested or one level clusters, so as to maximize the evaluation value of clustering where each cluster optimizes a clustering criterion [2, 5]. If *K* is the desired number of clusters, then partitioning approaches typically find all *K* clusters at once whereas traditional hierarchical schemes bisect a cluster to get two clusters or merge two clusters to get one at a time [5]. Hierarchical approach can be systematically used to generate a flat partition of *K* clusters and similarly the repeated application of a partitioning scheme can provide a hierarchical clustering [5]. Partitional clustering is of two types [6]:

- Hard Clustering
- Soft Clustering

In hard clustering, data is divided into different clusters such that each data item belongs to exactly a single cluster whereas in the case of soft clustering also called fuzzy clustering, data items can belong to more than one cluster. Each element has a set of membership levels that indicate the strength of the association between that element and a particular cluster. Fuzzy clustering is a process of allocating these membership levels and using them to assign data elements to one or more clusters. The Hard Approaches algorithm has a drawback that the cluster result is sensitive to the selection of the initial cluster centroids and may converge to the local optima and thus, the initial selection of the cluster centroids decides the local optimal solution in the vicinity of the initial solution of K-means and the partition result of the dataset [4]. The Soft Approaches algorithm is a population based stochastic optimization technique which is used to find an optimal or near optimal solution to a numerical and qualitative problem [4]. The Soft Approaches algorithm is used to generate good initial cluster centroids for optimal solution [4]. K-Means algorithm is an example of Hard Clustering approach while Fuzzy C-Means is an example of Soft Clustering approach.

III. Fuzzy C-Means Algorithm

Fuzzy C-Means was proposed by Dunn in 1973 and was modified by Bezdek in 1981. It is one of the most popular fuzzy clustering techniques with the approach that the data points have their membership values with the cluster centers that will be iteratively updated [3]. Fuzzy c-means clustering involves two major steps: the calculation of cluster centers and the assignment of points to these centers using a form of Euclidian distance such that the process is continuously repeated until the cluster centers stabilize [2]. The algorithm assigns a membership value to the data items for the clusters within a range of 0 to 1 and a fuzzification parameter in the range [1, n] which determines the degree of fuzziness in the clusters [2]. The FCM algorithm provides a method of clustering that enables a data item to belong to two or more clusters and this scheme of method is frequently used in pattern recognition applications [7]. It is based on minimization of the following objective function [3, 7, 10]:

$$J_{m_f} = \sum_{j=1}^{N}\sum_{k=1}^{C} \mu_{jk}^{m_f} \left\| x_j - c_k \right\|^2 \qquad (1)$$

Where:

m_f represents any real number greater than 1 such that $1 \leq m_f < \infty$, μ_{jk} is the degree of membership of x_j in the cluster j and c_k is the center of the cluster. In FCM, the membership matrix U is allowed to have not only 0 and 1 but also the elements with any values between 0 and 1, this matrix satisfying the following constraint [11]:

$$\sum_{j=1}^{C} \mu_{jk} = 1, \forall\, k = 1, \ldots, n \qquad (2)$$

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{jk} and the cluster centers c_k are given by [3, 7, 10]:

$$\mu_{jk} = \frac{1}{\sum_{p=1}^{C}\left[\frac{\left\|x_j - c_k\right\|}{\left\|x_j - c_p\right\|}\right]^{\frac{2}{m-1}}} \qquad (3)$$

$$c_k = \frac{\sum_{j=1}^{N} \mu_{jk}^{m_f} . x_k}{\sum_{j=1}^{N} \mu_{jk}^{m}} \qquad (4)$$

Fuzzy C-Means Algorithm Steps

The FCM algorithm consists of the steps as shown below [7, 10, 11]:

1. The membership matrix U is initialized with random values between 0 and 1 such that the constraints in Equation 2 are satisfied.
2. Calculate fuzzy cluster centers c_k , where k=1,..,C using Equation 3.
3. Calculate objective function according to Equation 1 and stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.
4. Compute a new membership matrix U using Equation 3.
5. Go to step 2.
6. This iteration will stop if
$$\left\| U^{(k+1)} - U^{k} \right\| < \xi, \quad (5)$$
where ξ is a termination criterion between 0 and 1, whereas k is the iteration steps.

Drawbacks of Fuzzy C-Means Algorithm

Fuzzy C-Means Algorithm suffers from certain drawbacks due to the restriction that the sum of membership values of a data point x_i in all the clusters must be equal to one as given by equation (4) [2]:

1. Firstly, this constraint tends to give high membership values for the outlier points and due to this the algorithm has difficulty in handling outlier points [2].
2. Secondly, in a cluster the membership of a data points depends directly on the membership values of other cluster centers which may lead to undesirable results [2].
3. FCM also faces problems in handling high dimensional data sets and a large number of prototypes. Also FCM is sensitive to initialization and is easily trapped in local optima [2].

IV. Possibilistic C-Means Algorithm (PCM)

As discussed in the previous section, FCM is the most popular and widely used clustering model algorithm. In FCM model let X= {x_1,x_2,x_3,…..,x_n} be a data set then the FCM algorithm assigns memberships to x_k which are inversely related to the relative distance of x_k to the c point prototypes denoting cluster centers such that if c=2 and if x_k is equidistant from the two prototypes then the membership of x_k in each cluster will be the same i.e. 0.5, regardless of the absolute value of the distance of x_k from the two centroids [12]. This leads to the problem of noise points that are equidistant from the two clusters and can be given equal membership in both when it is required that such points should be given very low or no membership in either of the cluster. To overcome this problem of noise points Krishnapuram and Keller proposed a new clustering algorithm known as possibilistic C-Means (PCM) which satisfies a relatively looser constraint i.e. each element of the i_{th} column can be any number between 0 and 1, as long as at least one of them is positive [12, 16]. The PCM algorithm considers the clustering problem from the viewpoint of possibility theory and its approach is different from the FCM algorithm because the resulting u_{jk} values can be represented in terms of degrees of possibility of the points belonging to the classes [13, 14]. The PCM algorithm helps to identify outliers (noise points). PCM algorithm results in low typicality values for outliers and automatically eliminates these noise points. At the same time PCM is very sensitive to initializations and may generate coincident clusters. In addition typicalities may be sensitive to the choice of the additional parameters needed by the PCM algorithm [12]. The objective function for PCM is given by [14, 15]:

$$P_m(T,V;X,\gamma) = \sum_{i=1}^{N}\sum_{k=1}^{C} t_{ik}^{m}\ d_{ki}^{2} + \sum_{i=1}^{C}\gamma_i \sum_{k=1}^{N}(1-t_{ki})^m \quad (6)$$

where:

t_{ki} is the typicality of x_k to the cluster i; vi; T is the typicality matrix, defined as T = $[t_{ki}]_{NC}$, d_{ki} is a distance measure between x_k and c_i and γ_i denotes a user-defined constant: $\gamma_i > 0$, $1 < i < c$, such that [14, 15]:

$$t_{ki} = 1\Big/\left(1+\frac{d_{ik}}{\gamma_i}\right)^{\frac{1}{m-1}}, \forall\, i,k \quad (7)$$

$$v_i = \frac{\sum_{k=1}^{n} t_{ki}^{m} x_k}{\sum_{k=1}^{n} t_{ki}^{m}}, \forall\, i \quad (8)$$

On solving equation 6 and 7next condition on γ_i can be derived as follows [14, 15]:

$$\gamma_i = K\,\frac{\sum_{k=1}^{N}\mu_{ki}^{m}\ d_{ki}^{2}}{\sum_{k=1}^{n}\mu_{ki}^{m}}\quad, K > 0 \quad (9)$$

Where μ_{ki} represents membership values and K=1in most cases. The PCM algorithm is more robust in the presence of noise and is efficient in finding valid cluster and thus gives a robust estimate of the centers [16]. In PCM updating of the membership values depends on the distance measurements namely, Euclidean distance that works effectively when a data set is compact or isolated and Mahalanobis distance takes into account the correlation in the data by using the inverse of the variance-covariance matrix of data set [16].

Advantages of PCM

1. PCM enables clustering of noisy data samples i.e. data sets with presence of outliers or noisy points [17].

Disadvantages of PCM

1. PCM is extremely sensitive to good initialization.
2. The algorithm may lead to generation of coincident clusters since the columns and rows of the typicality matrix are independent of each other. If the initialization of each row is not sufficiently distinct it may lead to coincident clusters [17, 12].

V. Fuzzy Possibilistic C-Means Algorithm (FPCM)

To overcome difficulties of the PCM, Pal and Bezdek, in 1997 proposed to integrate the features of both Fuzzy C-Means and Possibilistic C-Means by using the fuzzy values of the FCM as well as the typicality values of the PCM in order to achieve a better clustering model [14, 18]. They named this integrated approach as Fuzzy Possibilistic C-Means or FPCM. Membership and Typicality are very essential for the correct and precise characteristic of data substructure in clustering model and FPCM uses an objective function that depends on both membership and typicality features and is given as under [17, 18]:

$$J_{FPCM}(U,T,V) = \sum_{i=1}^{C}\sum_{j=1}^{N}(\mu_{ij}^{m} + t^{\eta})\ d^2\left(x_j, v_i\right) \qquad (10)$$

The algorithm also follows following constraints [18, 14]:

$$\sum_{i=1}^{C}\mu_{ij} = 1, \forall\, j\ \varepsilon\, \{1,\dots,n\} \qquad (11)$$

$$\sum_{i=1}^{C} t_{ij} = 1, \forall\, i\ \varepsilon\, \{1,\dots,c\} \qquad (12)$$

The first order necessary conditions for extreme of $J_{FPCM}(U,T,V)$ in terms of Lagrange multiplier theorem can be represented as follows [14]:

$$t_{ik} = 1\Big/\sum_{j=1}^{N}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/m-1}, \forall\, i,k \qquad (13)$$

$$v_i = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{m} + t_{ik}^{\eta}\right)x_k}{\sum_{k=1}^{N}\left(\mu_{ik} + t_{ik}^{\eta}\right)}, \forall\, i \qquad (14)$$

Where d_{ik} is the distance of the data point x_k to the prototype v_i, computed as [14, 18]:

$$d_{ik} = \|x_k - v_i\| = (x_k - v_i)^T A\,(x_k - v_i) \qquad (15)$$

Here, A is symmetric positive definite matrix. FPCM generates Memberships and possibilities at the same time, together with the usual point prototypes or cluster center for each cluster.

Advantages of FPCM [17]:

1) FPCM is a hybridization of possibilistic c-means (PCM) and fuzzy c-means (FCM) and avoids various problems of PCM and FCM.
2) FPCM enables to overcome the coincident clusters problem of PCM.
3) It also solves the noise sensitivity deficiency of FCM but the noisy data may have an influence on the estimation of centroids.

Disadvantages

1) The row sum constraints must be equal to one that may be problematic for big data sets [17, 18].

VI. Possibilistic Fuzzy C Means Algorithm (PFCM)

As discussed in the previous section in FPCM the constraint according to which the sum of all typicality values of all data to a cluster must be equal to one may lead to problems for big data sets [12, 15]. In order to solve this problem Pal [12] proposed a new and improved algorithm called Possibilistic Fuzzy c means algorithm (PFCM). The objective function for PFCM is given by the following equation [12, 17]:

$$\boldsymbol{J_{m,\eta}(U,T,V;Z) = \sum_{i=1}^{c}\sum_{k=1}^{n} a\mu_{ik}^{m} + bt_{ik}^{m} * \|z_k - v_i\|^2 + \sum_{i=1}^{c}\delta_i\sum_{k=1}^{n}(1 - t_{ik})^{\eta}} \qquad \mathbf{(16)}$$

The following constraints are followed [12, 17]:

$$\sum_{i=1}^{C}\sum_{k=1}^{n}\mu_{ik} = 1\ \forall k, 0 \le \mu_{ik}, t_{ik} \le 1, a > 0, b > 0, m > 1, \eta > 1 \qquad (17)$$

The parameters 'a' and 'b' define the relative importance between the membership degrees and the typicality values. As proposed, Theorem PFCM [12] defines the conditions to minimize the objective function $\boldsymbol{J_{m,\eta}}$. **The** theorem states that if $D_{ik} = \|x_k - v_i\|_A$ >0 for every i and k, m, η>1 and X contains at least c distinct data points then (U, T, V) ϵ $M_{FCM} \times M_{PCM} \times \Re^P$ minimizes objective function $\boldsymbol{J_{m,\eta}}$ only if [12, 17]:

$$\mu_{ik} = \left(\sum_{j=1}^{C}\left(\frac{D_{ikA}}{D_{jkA}}\right)^{\frac{2}{(m-1)}}\right)^{-1} \qquad (18)$$

$1 \le i \le c\,;\ 1 \le k \le \eta$

$$t_{ik} = \frac{1}{1 + \left(b\,(D_{ik}^2)/\delta_i\right)^{\frac{1}{(\eta-1)}}} \qquad (19)$$

$1 \le i \le c\,;\ 1 \le k \le \eta$

$$v_i = \frac{\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)x_k}{\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)} \qquad (20)$$

$1 \le i \le c$

Advantages of PFCM are as follows [12]:

1) PFCM solves the noise sensitivity deficiency problem of FCM algorithm.
2) PFCM helps to overcome the coincident clusters problem of PCM.
3) PFCM provides an improvement to FPCM by eliminating the row sum constraints of FPCM.

VII. Conclusion

In this paper a study of the four popular fuzzy clustering algorithms (FCM, PCM, FPCM and PFCM) has been presented. The Fuzzy clustering approaches overcome the drawbacks of the traditional clustering methods used earlier. FCM algorithm is the most popular fuzzy based clustering algorithm that has wide range of applications in different fields of study such as data mining applications, medicine, imaging, pattern detection, bioinformatics and other scientific and engineering applications. Moreover, various algorithms have been proposed and developed by many authors with Fuzzy C-Means algorithm as their basis and the goal of clustering more general datasets. In this study we have analyzed some of these algorithms like PCM, FPCM and PFCM each with their set of advantages and disadvantages. Careful review of these approaches suggests that further improvements in future by using advanced clustering techniques can help to achieve more accurate output and fast and efficient information retrieval even from larger dataset.

VIII. References

[1] Soumi Ghosh, Sanjay Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms", ((IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013

[2] Binsy Thomas, Madhu Nashipudimath," Comparative Analysis Of Fuzzy Clustering Algorithms In Data Mining" ISSN: 2277 – 9043 International J ournal of Advanced

Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 1, Issue 7, September 2012

[3] Subhagata Chattopadhyay, Dilip Kumar Pratihar, "A Comparative Study Of Fuzzy C-Means Algorithm And Entropy-Based Fuzzy Clustering Algorithms", Computing and Informatics, Vol. 30, 2011, 701–720.

[4] Vibekananda Dutta, Krishna Kumar Sharma, Deepti Gahalot, "Performance Comparison of Hard and Soft Approaches for Document Clustering", International Journal of Computer Applications (0975 – 8887) Volume 41– No.7, March 2012

[5] Michael Steinbach, George Karypis, Vipin Kumar, A "Comparison of Document Clustering Techniques" ,Department of Computer Science and Engineering, University of Minnesota, Technical Report #00-034 {steinbac, karypis, kumar@cs.umn.edu}

[6] Kamanashis Goswami, " A Study On Some Variations Of The Fuzzy C-Means Clustering Algorithm" Thesis Submitted Towards Partial Fulfillment Of The Requirements For The Award Of The Degree Of Master Of Electronics & Telecommunications Engineering, Department Of Electronics & Telecommunications Engineering Jadavpur University Kolkata-700032 India. May, 2010

[7] Dr. T. VELMURUGAN, "Performance Comparison between k-Means and Fuzzy C-Means Algorithms using Arbitrary Data Points", Vol 19, No. 8;Aug 2012 Wulfenia Journal

[8] Poonam fauzdar and Sujata Kindri , "Comparitive Analysis Of K Means And Fuzzy C Means Algorithm", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 2 Issue 6, June – 2013

[9] Md. Ehsanul Karim, Feng Yun, Sri Phani Venkata Siva Krishna Madani," Fuzzy Clustering Analysis", Thesis for the degree Master of Science in Mathematical Modelling and Simulation, Blekinge Institute of Technology School of Engineering Department of Mathematics and Science

[10] Pradeep Kumar Jena, Subhagata Chattopadhyay, " Comparative Study of Fuzzy k-Nearest Neighbor and Fuzzy C-means Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 57– No.7, November 2012

[11] Mofreh A. Hogo," Evaluation of E-Learners Behaviour using Different Fuzzy Clustering Models: A Comparative Study", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 2, 2010

[12] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek," A Possibilistic Fuzzy c-Means Clustering Algorithm", IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 13, NO. 4, AUGUST 2005

[13] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," IEEE Trans. Fuzzy Syst., vol. 1, no. 2, pp. 98–110, Apr. 1993.

[14] Christian Correa, Constantino Valero, Pilar Barreiro, Maria P. Diago, and Javier Tard_aguila," A Comparison of Fuzzy Clustering Algorithms Applied to Feature Extraction on Vineyard", http://www.etsia.upm.es

[15] B. Ojeda-Magaña , R. Ruelas , M. A. Corona-Nakamura , D. Andina ,"An Improvement To The Possibilistic Fuzzy C-Means Clustering Algorithm"

[16] M.H. Fazel Zarandi, M. Zarinbal, I.B. Turksen, "Type–II Fuzzy Possibilistic C-Mean Clustering" IFSA-EUSFLAT 2009

[17] R.Suganya, R.Shanthi," Fuzzy C- Means Algorithm- A Review", International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012 1 ISSN 2250-3153

[18] Mohamed Fadhel Saad and Adel M. Alimi, "Modified Fuzzy Possibilistic C-means", Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong